

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開2003-30240

(P2003-30240A)

(43)公開日 平成15年1月31日(2003.1.31)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード* (参考)
G 0 6 F 17/30	3 6 0	G 0 6 F 17/30	3 6 0 Z 5 B 0 7 5
	2 1 0		2 1 0 A 5 B 0 8 2
	3 5 0		2 1 0 D
12/00	5 1 3	12/00	3 5 0 C
			5 1 3 D
審査請求 有 請求項の数9 O L (全 8 頁)			

(21)出願番号 特願2001-214354(P2001-214354)

(22)出願日 平成13年7月13日(2001.7.13)

(71)出願人 000102728

株式会社エヌ・ティ・ティ・データ

東京都江東区豊洲三丁目3番3号

(72)発明者 末永 高志

東京都江東区豊洲三丁目3番3号 株式会

社エヌ・ティ・ティ・データ内

(72)発明者 坂野 鋭

東京都江東区豊洲三丁目3番3号 株式会

社エヌ・ティ・ティ・データ内

(74)代理人 100064908

弁理士 志賀 正武 (外2名)

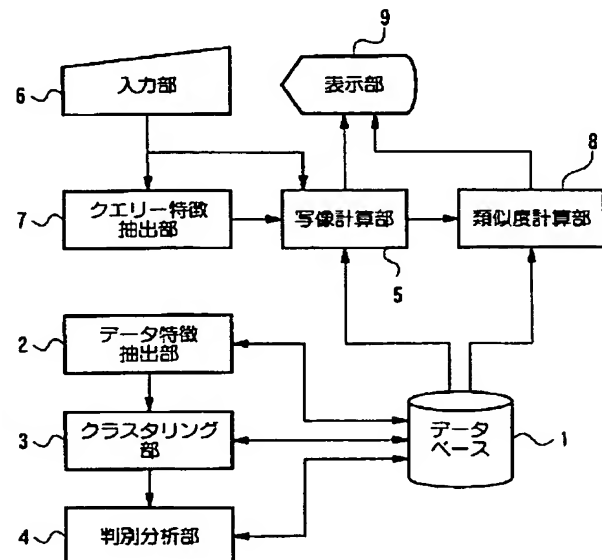
最終頁に続く

(54)【発明の名称】 データ検索装置、データ検索方法、及びデータ検索プログラム

(57)【要約】

【課題】 データベースに保存されているデータの相互関係がわかりやすい形で検索者に対して提示することが可能なデータ検索装置を提供する。

【解決手段】 データベースに保存されている全てのデータのそれぞれから4次元以上の次元数を有する特徴量を抽出する手段と、データの特徴量に基づいて、データベースに保存されている複数のデータを所定の数 of クラスタに分ける手段と、クラスタ分けがされたデータに対して判別分析を使用して各データの特徴量の次元数を3次元以下にするための射影行列を算出する手段と、入力されたクエリーから4次元以上の次元数を有する特徴量を抽出する手段と、データの特徴量及びクエリーの特徴量に対して射影行列を乗算して次元数が3次元以下となる座標値を求め、この座標値をプロットすることにより、データベースに保存されている各データとクエリーの関係を散布図によって表示する手段とを備える。



## 【特許請求の範囲】

【請求項1】 データベースに保存されている複数のデータの中から所望のデータを検索するデータベース検索装置であって、

前記データベース検索装置は、

前記データベースに保存されている全てのデータのそれぞれから4次元以上の次元数を有する特徴量を抽出するデータ特徴抽出手段と、

前記データの特徴量に基づいて、前記データベースに保存されている複数のデータを所定の数のクラスタに分けるクラスタリング手段と、

前記クラスタリング手段によってクラスタ分けがされたデータに対して判別分析を使用して各データの特徴量の次元数を3次元以下にするための射影行列を算出する射影行列算出手段と、

所望のデータを検索するために入力されたクエリーから4次元以上の次元数を有する特徴量を抽出するクエリー特徴抽出手段と、

前記データの特徴量及び前記クエリーの特徴量に対して前記射影行列を乗算して次元数が3次元以下となる座標値を求め、この座標値をプロットすることにより、前記データベースに保存されている各データとクエリーの関係を散布図によって表示する写像計算手段と、  
を備えたことを特徴とするデータ検索装置。

【請求項2】 前記データ検索装置は、前記写像計算手段によってプロットされたクエリーの特徴量の近傍に位置するデータを選択するデータ選択手段と、

前記データ選択手段によって選択されたデータのそれぞれとクエリーの類似度を4次元以上の特徴量に基づいて計算して表示する類似度計算手段と、  
をさらに備えたことを特徴とする請求項1に記載のデータ検索装置。

【請求項3】 前記類似度計算手段は、4次元以上の次元数を有する特徴量のユークリッド距離を類似度とすることを特徴とする請求項2に記載のデータ検索装置。

【請求項4】 データベースに保存されている複数のデータの中から所望のデータを検索するデータベース検索方法であって、

前記データベース検索方法は、

前記データベースに保存されている全てのデータのそれぞれから4次元以上の次元数を有する特徴量を抽出するデータ特徴抽出過程と、

前記データの特徴量に基づいて、前記データベースに保存されている複数のデータを所定の数のクラスタに分けるクラスタリング過程と、

前記クラスタリング過程によってクラスタ分けがされたデータに対して判別分析を使用して各データの特徴量の次元数を3次元以下にするための射影行列を算出する射

影行列算出過程と、

所望のデータを検索するために入力されたクエリーから4次元以上の次元数を有する特徴量を抽出するクエリー特徴抽出過程と、

前記データの特徴量及び前記クエリーの特徴量に対して前記射影行列を乗算して次元数が3次元以下となる座標値を求め、この座標値をプロットすることにより、前記データベースに保存されている各データとクエリーの関係を散布図によって表示する写像計算過程と、  
を有することを特徴とするデータ検索方法。

【請求項5】 前記データ検索方法は、前記写像計算過程によってプロットされたクエリーの特徴量の近傍に位置するデータを選択するデータ選択過程と、

前記データ選択過程によって選択されたデータのそれぞれとクエリーの類似度を4次元以上の特徴量に基づいて計算して表示する類似度計算過程と、  
をさらに有することを特徴とする請求項4に記載のデータ検索方法。

【請求項6】 前記類似度計算過程は、4次元以上の次元数を有する特徴量のユークリッド距離を類似度とすることを特徴とする請求項5に記載のデータ検索方法。

【請求項7】 データベースに保存されている複数のデータの中から所望のデータを検索するデータベース検索プログラムであって、

前記データベース検索プログラムは、

前記データベースに保存されている全てのデータのそれぞれから4次元以上の次元数を有する特徴量を抽出するデータ特徴抽出処理と、

前記データの特徴量に基づいて、前記データベースに保存されている複数のデータを所定の数のクラスタに分けるクラスタリング処理と、

前記クラスタリング処理によってクラスタ分けがされたデータに対して判別分析を使用して各データの特徴量の次元数を3次元以下にするための射影行列を算出する射影行列算出処理と、

所望のデータを検索するために入力されたクエリーから4次元以上の次元数を有する特徴量を抽出するクエリー特徴抽出処理と、

前記データの特徴量及び前記クエリーの特徴量に対して前記射影行列を乗算して次元数が3次元以下となる座標値を求め、この座標値をプロットすることにより、前記データベースに保存されている各データとクエリーの関係を散布図によって表示する写像計算処理と、  
をコンピュータに行わせることを特徴とするデータ検索プログラム。

【請求項8】 前記データ検索プログラムは、前記写像計算処理によってプロットされたクエリーの特徴量の近傍に位置するデータを選択するデータ選択処理

と、  
前記データ選択処理によって選択されたデータのそれぞれとクエリーの類似度を4次元以上の特徴量に基づいて計算して表示する類似度計算処理と、  
をさらにコンピュータに行わせることを特徴とする請求項7に記載のデータ検索プログラム。

【請求項9】 前記類似度計算処理は、  
4次元以上の次元数を有する特徴量のユークリッド距離を類似度とすることを特徴とする請求項8に記載のデータ検索プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、大規模データベースにおける検索結果の表示をユーザに理解しやすい形で提示するデータ検索装置、データ検索方法、及びデータ検索プログラムに関する。

【0002】

【従来の技術】従来、データベースからのデータ検索においては入力された質問（クエリー）に対して最も類似したデータ群が類似性の順番で出力されることが一般的である。この方法によるデータ検索は、検索者の意図と異なったデータが大量に出力されることが多いため、検索者は、さらに絞り込むための質問を入力するか、大量のデータの中から所望のデータを探し出さなければならず、検索者の負担が大きく効率が悪いという問題がある。

【0003】このような問題を解決するために、データベースに保存されたデータを複数の数値で構成される特徴ベクトル表現し、保存されているデータの関係を人間が把握可能な2もしくは3次元の散布図の形で表示することによりデータ同士の関係を空間的に表示する方法が研究されている。この場合、データを表現するベクトルの次元数が4以上の場合、散布図に表示するために2もしくは3次元の座標を求める次元削減技術が要求されるが、従来はこの次元削減のために主成分分析が用いられてきた。

【0004】

【発明が解決しようとする課題】しかしながら、多くのデータベースではデータを表現するために必要な情報が多いため、データを表現する特徴ベクトルは少なくとも100次元、多いときには数千次元のベクトルとなる。従来の主成分分析を用いた表示では、データ数が10数個程度の、すなわちデータが低次元の部分空間に縮重している小規模データベースでは検索のために効果的な散布図が生成されるが、データ数が数百を超えると、意味のある表示が得られることはまれで、本来の目的である効率的な検索を行うことが困難であるという問題がある。また、主成分分析を用いたのでは、高次元空間に存在する数千から数万の特徴ベクトルについて、その全ての距離関係を保存することを目的として低次元への写像

を行うため、全ての関係が少しずつ損なわれ、結果として低次元に写像されたデータの分布は特徴空間での分布構造をほとんどもしくはまったく反映しない表示となる。すなわち、主成分分析を用いる限り、特に大規模データベースを対象にした場合には特徴空間での遠近関係を低次元空間に保存する表示を得ることはできないため、結果として検索効率を高くすることはできないという問題がある。

【0005】本発明は、このような事情に鑑みてなされたもので、データベースに保存されているデータの相互関係がわかりやすい形で検索者に対して提示することが可能なデータ検索装置、データ検索方法、及びデータ検索プログラムを提供することを目的とする。

【0006】

【課題を解決するための手段】請求項1に記載の発明は、データベースに保存されている複数のデータの中から所望のデータを検索するデータベース検索装置であって、前記データベース検索装置は、前記データベースに保存されている全てのデータのそれぞれから4次元以上の次元数を有する特徴量を抽出するデータ特徴抽出手段と、前記データの特徴量に基づいて、前記データベースに保存されている複数のデータを所定の数のクラスターに分けるクラスタリング手段と、前記クラスタリング手段によってクラスター分けがされたデータに対して判別分析を使用して各データの特徴量の次元数を3次元以下にするための射影行列を算出する射影行列算出手段と、所望のデータを検索するために入力されたクエリーから4次元以上の次元数を有する特徴量を抽出するクエリー特徴抽出手段と、前記データの特徴量及び前記クエリーの特徴量に対して前記射影行列を乗算して次元数が3次元以下となる座標値を求め、この座標値をプロットすることにより、前記データベースに保存されている各データとクエリーの関係を散布図によって表示する写像計算手段とを備えたことを特徴とする。

【0007】請求項2に記載の発明は、前記データ検索装置は、前記写像計算手段によってプロットされたクエリーの特徴量の近傍に位置するデータを選択するデータ選択手段と、前記データ選択手段によって選択されたデータのそれぞれとクエリーの類似度を4次元以上の特徴量に基づいて計算して表示する類似度計算手段とをさらに備えたことを特徴とする。

【0008】請求項3に記載の発明は、前記類似度計算手段は、4次元以上の次元数を有する特徴量のユークリッド距離を類似度とすることを特徴とする。

【0009】請求項4に記載の発明は、データベースに保存されている複数のデータの中から所望のデータを検索するデータベース検索方法であって、前記データベース検索方法は、前記データベースに保存されている全てのデータのそれぞれから4次元以上の次元数を有する特徴量を抽出するデータ特徴抽出過程と、前記データの特

微量に基づいて、前記データベースに保存されている複数のデータを所定の数のクラスタに分けるクラスタリング過程と、前記クラスタリング過程によってクラスタ分けがされたデータに対して判別分析を使用して各データの特徴量の次元数を3次元以下にするための射影行列を算出する射影行列算出過程と、所望のデータを検索するために入力されたクエリーから4次元以上の次元数を有する特徴量を抽出するクエリー特徴抽出過程と、前記データの特徴量及び前記クエリーの特徴量に対して前記射影行列を乗算して次元数が3次元以下となる座標値を求め、この座標値をプロットすることにより、前記データベースに保存されている各データとクエリーの関係を散布図によって表示する写像計算過程とを有することを特徴とする。

【0010】請求項5に記載の発明は、前記データ検索方法は、前記写像計算過程によってプロットされたクエリーの特徴量の近傍に位置するデータを選択するデータ選択過程と、前記データ選択過程によって選択されたデータのそれぞれとクエリーの類似度を4次元以上の特徴量に基づいて計算して表示する類似度計算過程とをさらに有することを特徴とする。

【0011】請求項6に記載の発明は、前記類似度計算過程は、4次元以上の次元数を有する特徴量のユークリッド距離を類似度とすることを特徴とする。

【0012】請求項7に記載の発明は、データベースに保存されている複数のデータの中から所望のデータを検索するデータベース検索プログラムであって、前記データベース検索プログラムは、前記データベースに保存されている全てのデータのそれぞれから4次元以上の次元数を有する特徴量を抽出するデータ特徴抽出処理と、前記データの特徴量に基づいて、前記データベースに保存されている複数のデータを所定の数のクラスタに分けるクラスタリング処理と、前記クラスタリング処理によってクラスタ分けがされたデータに対して判別分析を使用して各データの特徴量の次元数を3次元以下にするための射影行列を算出する射影行列算出処理と、所望のデータを検索するために入力されたクエリーから4次元以上の次元数を有する特徴量を抽出するクエリー特徴抽出処理と、前記データの特徴量及び前記クエリーの特徴量に対して前記射影行列を乗算して次元数が3次元以下となる座標値を求め、この座標値をプロットすることにより、前記データベースに保存されている各データとクエリーの関係を散布図によって表示する写像計算処理とをコンピュータに行わせることを特徴とする。

【0013】請求項8に記載の発明は、前記データ検索プログラムは、前記写像計算処理によってプロットされたクエリーの特徴量の近傍に位置するデータを選択するデータ選択処理と、前記データ選択処理によって選択されたデータのそれぞれとクエリーの類似度を4次元以上の特徴量に基づいて計算して表示する類似度計算処理と

をさらにコンピュータに行わせることを特徴とする。

【0014】請求項9に記載の発明は、前記類似度計算処理は、4次元以上の次元数を有する特徴量のユークリッド距離を類似度とすることを特徴とする。

【0015】

【発明の実施の形態】以下、本発明の一実施形態によるデータ検索装置を図面を参照して説明する。図1は同実施形態の構成を示すブロック図である。この図において、符号1は、検索対象の大規模データベースであり、ここでは、文書データが保存されているものとする。符号2は、データベース1に保存されている各データを高次元数値ベクトルデータに変換することによってデータの特徴を抽出するデータ特徴抽出部である。符号3は、データベース1に保存されている高次元数値ベクトルデータに対してクラスタリングを行うクラスタリング部である。符号4は、クラスタリングされた高次元数値ベクトルデータに対して判別分析を行う判別分析部である。符号5は、判別分析によって得られた写像を用いてデータを表す高次元ベクトルを低次元に写像する写像計算部である。符号6は、質問（以下、クエリーという）を入力するクエリー入力部であり、キーボード等で構成される。なお、入力部6は、データファイルを読み込む構成であってもよい。符号7は、入力部6より入力されたクエリーの特徴を抽出するクエリー特徴抽出部である。符号8は、クエリーとデータとの類似度を計算する類似度計算部である。符号9は、CRT、液晶ディスプレイなどで構成される表示部である。

【0016】ここで、本発明のデータ検索の原理を図5、6を参照して、簡単に説明する。本発明は、類似するデータの集合を人間が直感的に認識しやすいように、高次元の数値ベクトルを2次元表現するための次元削減することが目的である。ここでは、簡単のために3次元の数値ベクトルを2次元表現することを例にして説明する。図5の(a)は、各データの特徴ベクトルを点で表現した図である。この図において、距離が近い点は類似しているデータであると見なし、k平均法を用いて各データをクラスタ分けする。そして、図5(a)の破線で示す2次元平面を判別分析によって求め、この2次元平面に各点を写像すると、図5の(c)に示すような図が得られる。この2次元平面に対して検索条件に相当するクエリーをプロットすれば、条件に近いデータの集まりを直感的に知ることができる。

【0017】一方、写像をとる2次元平面が適切でない場合は、図5の(b)のように、類似しているデータの集まりを認識することが困難である。本発明は、データが高次元の数値ベクトルである場合において、次元の特徴量が極力失われず、かつ類似データの集まりを直感的に認識することができる2次元平面を効率よく求めることが目的である。このために、本発明は、図6に示すように各データをクラスタ分けした後、クラスタ中心の2

次元平面の写像点と各データの分散値が均等になる2次元平面を、判別分析方を用いて求めるものである。このようにすることによって、あるクラスに属するデータを近くに集め、かつ異なるクラスに属するデータを分離して、データ検索者に対して表示することが可能となる。

【0018】次に、図1に示すデータ検索装置の検索動作を説明する。ここでは、ヘルプデスクでの問い合わせ電子メールに対し、過去に類似の問い合わせが無かったかなどを検索する目的で、データベース1に保存されている文書から類似文書を検索する動作を例にして説明する。初めに、データ検索を行う前のオフライン処理について説明する。まず、データ特徴抽出部2は、データベース1に保存されている文書データを読み出す(ステップS1)。そして、読み出した文書データの特徴ベクトル $x_n$  ( $n=1 \cdots N$ )を求める(ステップS2)。この特徴ベクトル $x_n$ は、データ検索に必要な複数の単語のそれぞれの出現頻度のヒストグラムに基づいて求められ、データベース1に文書データに関連付けて保存する。

【0019】例えば、予め決められているデータ検索に必要な単語が、「コンピュータ」、「連携」、「ディスプレイ」、「キーボード」と定義されており、対象の文書データが、「コンピュータはコンピュータ自身のみではなく周囲の装置と連携することにより動作する」という場合、「コンピュータ」が2個、「連携」が1個、「ディスプレイ」及び「キーボード」がそれぞれ0個であるので、この文書は(2、1、0、0)といった複数の数値の組み合わせ、すなわちベクトルとして表現され、文書データと関連付けられて保存される。この処理は、データベース1に保存されている全ての文書データに対して実行され、この時点で、データベース1には、文書データ毎に特徴ベクトルが関連付けられて保存された状態となる。続いて、データ特徴抽出部2は、クラスタリング部3に対して特徴抽出が終了したことを通知する。

【0020】次に、クラスタリング部3は、データベース1に保存されている文書データをランダムに $k$  ( $k$ は2以上の自然数)個取り出し、この $k$ 個の文書データを仮のクラス中心とし(ステップS3)、それぞれに1〜 $k$ のクラス番号を付与する。続いてクラスタリング部3は、データベース1に保存されている文書データを順に読み出す。そして、先に取り出した $k$ 個の文書データの中から、読み出した文書データが最も近い文書データを求め、読み出した文書データに対して最も近い文書データのクラス番号を仮に付与する。ここでいう最も近いとは、特徴ベクトルのユークリッド距離が最も近いことである。この処理を全ての文書データに対して施す。これによって、全ての文書データに対して、1〜 $k$ のいずれかのクラス番号が仮に付与され、文書データ

は $k$ 個のクラスに分類されたこととなる。

【0021】次に、クラスタリング部3は、各クラスに属する文書データの部分集合の平均値を求め、この平均値を新たなクラス中心とする(ステップS5)。そして、クラスタリング部3は、新たなクラス中心が直前のクラス中心と同じになるまでステップS4、S5の処理を繰り返す(ステップS6)、各文書データに対して仮に付与したクラス番号をラベルとして文書データに追加して、データベース1に保存する(ステップS7)。続いてクラスタリング部3は、クラスタリング処理が終了したことを判別分析部4へ通知する。

【0022】次に、判別分析部4は、データベース1に保存されている $N$ 個の文書データの特徴ベクトルの全平均 $m$ を計算する(ステップS8)。続いて、判別分析部4は、各クラス1〜 $k$ の平均 $m_i$ を計算する(ステップS9)。そして、判別分析部4は、クラス内分散行列 $S_w$ 及びクラス間分散行列 $S_b$ を計算する(ステップS10、S11)。そして、判別分析部4は、 $S_w^{-1} S_b$ の固有値問題を解く(ステップS12)。すなわち、各クラスの距離が遠くなり、かつクラス内の各データの距離が近くなる解を求める。

【0023】次に、判別分析部4は、ステップS12において得られた固有値を大きい順に並び替え(ステップS13)、1番目及び2番目の固有値に対応する固有ベクトル $W$ を取り出す(ステップS14)。そして、データベース1に保存されている全ての文書データに対して行列演算することにより座標 $y_n$ を計算して(ステップS15)、その結果をデータベース1に保存する。このように、ステップS1〜S15のオフライン処理によって、データベース1に保存されている文書データが $k$ 個のクラスに分けられ、かつ高次元のベクトルデータが2次元表現可能なデータである座標 $y_n$ に変換されたこととなり、この座標 $y_n$ をプロットすれば散布図が得られる。なお、図2に示すステップS1〜S15のオフライン処理は、データベース1に新たに文書データが保存されるのに合わせて、定期的に行われる。

【0024】次に、前述したオフライン処理が施された文書データから所望のデータを検索する動作を説明する。まず、オペレータはヘルプデスクにメールが届くと、このメールをクエリーとして入力する(ステップS21)。入力部6は、このメールの内容を読み取り、その内容をクエリー特徴抽出部7へ出力する。これを受けて、クエリー特徴抽出部7はメールの内容を単語に分割し、その単語の出現頻度によって数値化し、特徴ベクトル $u$ を求め(ステップS22)、この特徴ベクトル $u$ を写像計算部5へ出力する。

【0025】次に、写像計算部5は、前述したステップS14において求めた射影行列(固有ベクトル) $W$ を用いてクエリーの座標 $v$ を計算する(ステップS23)。そして、写像計算部5は、求めた座標 $v$ を表示部9へ表

示する。また、写像計算部5は、データベース1に保存されている散布図のデータ（ステップS15において求めた座標 $y_n$ ）を読み出して、クエリーの座標 $v$ を表示した画面に重ね合わせて表示する（ステップS24）。この画面を見て、オペレータは、クエリーに近いデータを検索対象として入力部6によって選択する。この時点で表示部9に表示される散布図の一例を図4に示す。図4は、予め定義された単語が2000語、データベースに保存された文書データが500個、 $k$ の値が6である場合の処理実行例である。図4において、黒点が座標 $y_n$ に基づいてプロットされた各データを表し、符号Qは座標 $v$ に基づいてプロットされたクエリーを表している。また、符号Aは、オペレータが選択したエリアを示している。

【0026】写像計算部5は、このエリア内に存在するデータが、クエリーに類似したデータであると見なし、このエリア内に存在するデータを類似度計算部8へ通知する。これを受けて、類似度計算部8は、写像計算部5から通知されたデータのみを検索対象として、類似度の計算を行い、その結果を表示部9へ表示する。ここでいう類似度とは、ステップS2において求めた高次元の特徴ベクトルのユークリッド距離の近い順に類似度が高いと見なしたものである。この類似度が高い文書データを選択すれば、クエリーのメールの内容に近い文書データを探し出すことができる。

【0027】このように、データベース1に保存されているデータとクエリーの間を散布図によって表現するようにしたため、検索者はクエリーの近傍に位置するデータに注目することによって検索の効率を向上させることが可能となる。これはパターン認識辞書の設計補助や、文字認識、音声認識にも応用可能であり、さらに、CRM（Customer Relationship Management）などで用いられるデータマイニング技術への応用も可能である。

【0028】なお、前述した説明においては、クラスタリングを行う方法として、 $k$ 平均法を例にして説明したが、Word法などのクラスタリング手法を用いることも可能である。また、図4においては、データの関係を2次元表現するようにしたが、データの関係を3次元表現によって表現するようにしてもよい。このとき、ステップS14において、1～3番目の固有値に対する固有ベクトル $W = (w_1, w_2, w_3)$ を取り出すようにすればよい。

【0029】なお、図2、3に示す各処理の機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することによりデータ検索処理を行ってもよい。なお、ここでいう「コンピュータシステム」とは、OSや周辺機器等のハードウェアを含むものとする。また、「コンピュータ

システム」は、WWWシステムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムが送信された場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリ（RAM）のように、一定時間プログラムを保持しているものも含むものとする。

【0030】また、上記プログラムは、このプログラムを記憶装置等に格納したコンピュータシステムから、伝送媒体を介して、あるいは、伝送媒体中の伝送波により他のコンピュータシステムに伝送されてもよい。ここで、プログラムを伝送する「伝送媒体」は、インターネット等のネットワーク（通信網）や電話回線等の通信回線（通信線）のように情報を伝送する機能を有する媒体のことをいう。また、上記プログラムは、前述した機能の一部を実現するためのものであってもよい。さらに、前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるもの、いわゆる差分ファイル（差分プログラム）であってもよい。

【0031】

【発明の効果】以上説明したように、この発明によれば、高次元ベクトルとして表現されたデータの空間的関係を保存したまま人間に理解可能な低次元空間に写像することが可能になるため、データベースの検索効率が向上するという効果が得られる。

【図面の簡単な説明】

【図1】 本発明の一実施形態の構成を示すブロック図である。

【図2】 図1に示すデータ検索装置の動作を示すフローチャートである。

【図3】 図1に示すデータ検索装置の動作を示すフローチャートである。

【図4】 表示部9に表示される画面の一例を示す説明図である。

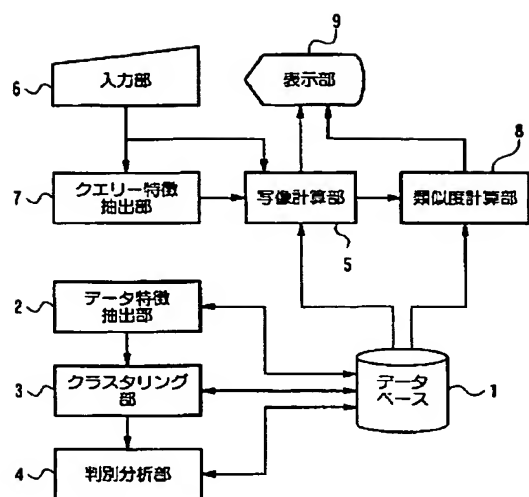
【図5】 本発明によるデータ検索原理を説明するための説明図である。

【図6】 本発明によるデータ検索原理を説明するための説明図である。

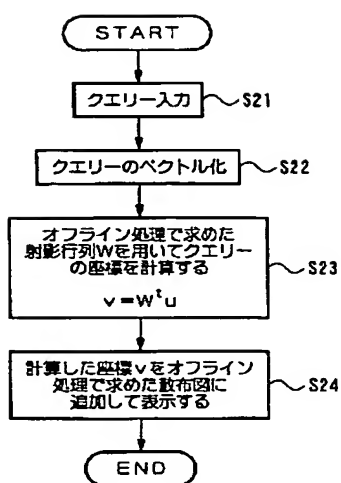
【符号の説明】

1・・・データベース、 2・・・データ特徴抽出部、 3・・・クラスタリング部、 4・・・判別分析部、 5・・・写像計算部、 6・・・入力部、 7・・・クエリー特徴抽出部、 8・・・類似度計算部、 9・・・表示部。

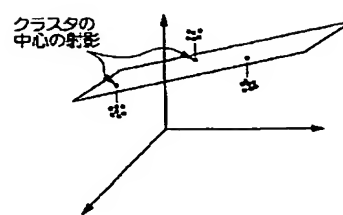
【図1】



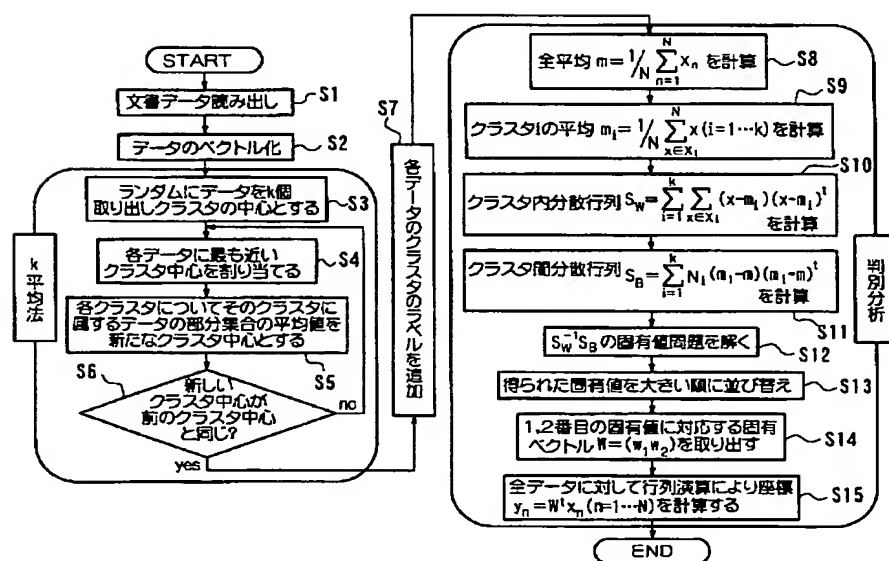
【図3】



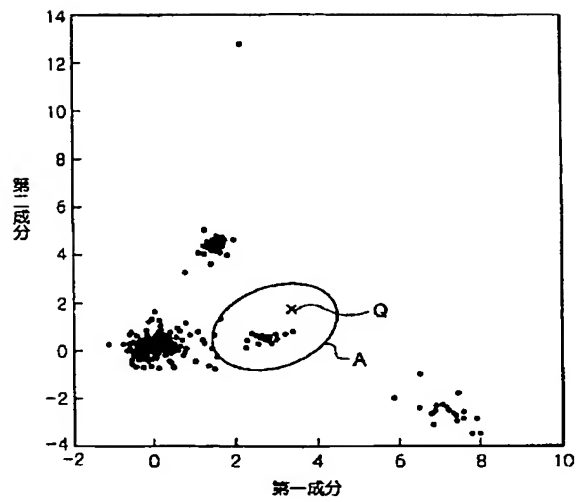
【図6】



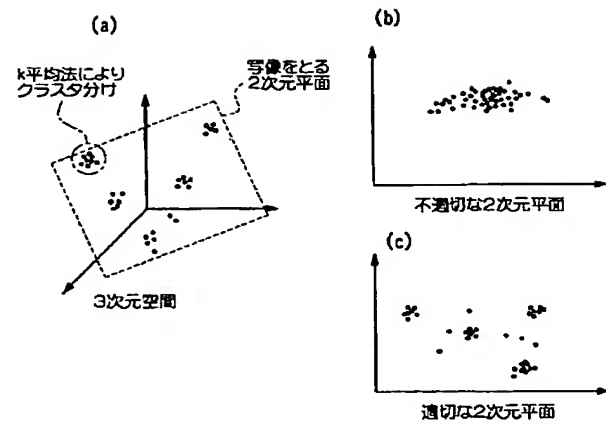
【図2】



【図4】



【図5】



フロントページの続き

F ターム(参考) 5B075 NK06 NR12 PQ02 PQ14 PRO6  
 QM08 QR02  
 5B082 GA08